

# THE PAIRING OF ACCRETING MASSIVE BLACK HOLES IN MULTIPHASE CIRCUMNUCLEAR DISKS: THE INTERPLAY BETWEEN RADIATIVE COOLING, STAR FORMATION, AND FEEDBACK PROCESSES

RAFAEL SOUZA LIMA<sup>1</sup>, LUCIO MAYER<sup>1</sup>, PEDRO R. CAPELO<sup>1</sup>, AND JILLIAN M. BELLOVARY<sup>2</sup>

<sup>1</sup>Center for Theoretical Astrophysics and Cosmology, Institute for Computational Science, University of Zurich, Winterthurerstrasse 190, CH-8057 Zürich, Switzerland

<sup>2</sup>Department of Astrophysics, American Museum of Natural History, Central Park West & 79th St., New York, NY 10024, USA

## ABSTRACT

We study the orbital decay of a pair of massive black holes (BHs), in the mass range  $5 \times 10^5 - 10^7 M_\odot$ , using a large set of hydrodynamical simulations of circumnuclear disks (CNDs) with varying prescriptions for the sub-grid physics of the interstellar medium, from star formation and supernova feedback to BH accretion and its feedback. In the absence of any of such processes, the orbit of the secondary BH in an adiabatic flow decays over timescales of a few Myr to the center of the CND, where the primary BH resides. As soon as strong dissipation operates in CNDs, fragmentation into massive objects the size of giant molecular clouds occurs, causing stochastic torques as well as direct hits that eject the secondary BH out of the disk plane. Once outside the plane, the low-density medium provides only weak drag, and the return to the CND plane is governed by inefficient dynamical friction in a stellar bulge. Ejections are seen to occur in nearly all of runs with cooling, irrespective of which other processes are modeled. In rare cases, clump–BH interactions can lead to a speed-up of the decay. Feedback processes lead to outflows but do not change significantly the overall density of the CND midplane, thus having minor direct impact on the orbital decay. However, with BH feedback a hot bubble is generated behind the secondary already on the first orbit, which almost shuts off dynamical friction, a phenomenon that we dub “wake evacuation”. In the latter case we find that delays in the decay of up to a fraction of Gyr can occur. We discuss the non-trivial implications on the discovery space of the evolved Laser Interferometer Space Antenna (eLISA). Our results suggest that the largest uncertainty in predicting massive BH merger rates lies in the potentially wide variety of galaxy host systems, with different degrees of gas dissipation and heating, yielding decay timescales from as small as 10 Myr to as large as a fraction of Gyr.

*Keywords:* black hole physics – galaxies: nuclei – Hydrodynamics – Methods: numerical

## 1. INTRODUCTION

In a wide range of galaxy merger remnants, two or more massive black holes (BHs) are expected. This is mainly due to two reasons. First, there is compelling evidence for the existence of massive BHs in the center of massive galaxies (Richstone et al. 1998; Magorrian et al. 1998). Second, in the  $\Lambda$ CDM model, the assembly of larger galaxies happens through hierarchical merging (White & Rees 1978). Therefore, if more than one galaxy undergoing a given merging process hosts a BH, then two or more BHs are expected in the merger remnant. Previous studies of BH pair evolution in galaxy mergers suggest that BHs are driven to the center of the new host galaxy (Begelman et al. 1980), by losing orbital energy and angular momentum via processes such as dynamical friction against the gaseous and stellar background (Ostriker 1999; Chandrasekhar

1943; Binney & Tremaine 2008). After the galaxies merge, the BHs reach a separation of less than  $\sim 10$  pc in relatively short timescales, compared to those in which the original galaxies merged (Callegari et al. 2009; Van Wassenhove et al. 2014; Capelo et al. 2015).

At a given stage of a gas-rich galaxy merger, the BH pair is expected to reside in a gaseous circumnuclear disk (CND). The formation of CNDs is predicted by simulations of mergers (Mayer et al. 2007; Mayer et al. 2010), furthermore it is corroborated by observations of nearby merging galaxies and galaxies that have undergone recent mergers (Medling et al. 2014). The mechanisms that compress and drive gas inflows to the nuclear regions of a merger (Mayer et al. 2007; Mayer et al. 2010; Capelo & Dotti 2016) likely also induce star formation in some CNDs (Downes & Solomon 1998). This idea is supported by several reports of massive star formation activity in the nuclei of galaxy mergers. CNDs with strong star formation activity are expected to be clumpy (Klessen et al. 2010; Agertz et al. 2009; Tasker & Tan 2009), since sites of star formation are gravitationally

unstable regions of gas, which can be associated to relatively cold and overdense gas clouds.

Recent studies on BH pair dynamics in CNDs attempted to track the influence of different interstellar medium properties, such as density contrasts (smooth vs. inhomogeneous), star formation, and gas accretion onto the BHs (Escala et al. 2005; Dotti et al. 2007; Mayer 2013; Fiacconi et al. 2013; Roškar et al. 2015; del Valle et al. 2015; Lupi et al. 2015). The effects of a clumpy environment on the dynamics of a BH pair in a range of BH mass ratios (1:1 to 1:5) in a gaseous CND are studied in Fiacconi et al. (2013). The “lumpiness” of the disk is triggered and tuned by considering an artificial cooling term in the energy equation satisfied by the gas particles. It was found that the orbit of the light secondary BH can be affected in a stochastic manner, often extending the decay time in comparison to the smooth disk scenario.

A similar result was obtained in Roškar et al. (2015) while considering an extended set of physical processes, however in galactic-scale simulations. According to this study, these processes affect the nature of the ambient in which the orbital decay process takes place. An implementation of a cooling term that accounts for energy-dissipation rates derived in atomic physics was considered, in contrast to the work of Fiacconi et al. (2013).

In the present work, the evolution of BH pairs in CNDs is studied. The effects of gas cooling, star formation, and SN feedback on the interstellar medium are included, using sub-grid prescriptions drawn from realistic simulations of galaxy formation that were not employed yet in simulations of CNDs (the gravitational force being resolved down to 0.5 pc), except in the work of Roškar et al. (2015). For the first time, the effects of BH accretion and feedback from active galactic nuclei (AGN) in simulations of CNDs are also considered. This study represents a complementary step in the exploration initiated by Roškar et al. (2015), by considering a large suite of CND runs rather than only a couple of galaxy merger runs. In this way we aim to understand in detail, and systematically, the effects of the various physical processes on the orbital decay. The BH masses are chosen to be in the range of those relevant for the future evolved Laser Interferometer Space Antenna (eLISA). Concerning other previous similar works, Fiacconi et al. (2013) carried out a large set of CND simulations with massive BH pairs, but did not consider the effects of star formation and SNe, nor BH accretion or AGN feedback. del Valle et al. (2015), in their implementation of SN feedback, let the gas cool after deposition of the energy of SN explosions into the surrounding gas, which is known to lead to inefficient feedback, likely missing the effect of blastwaves. They also did not include BH accretion nor its feedback.

We thus ran a comprehensive set of simulations to disentangle the effects of different processes on the dynamics of a BH pair in a CND. The set up of the initial conditions of these simulations is described in §2, in addition to the implementations of the physics considered. §3 highlights the results of the individual influence of gas cooling, star formation rates, SN feedback and AGN feedback on the orbital decay of the BH pair. Finally, in §4 we discuss the implications of the results on the current understanding of the process of orbital decay and on detection rates of future gravitational-wave experiments, the conclusions that can be drawn from them and limitations of the present work.

## 2. METHODS

Numerical simulations are performed using the Tree Smoothed Particle Hydrodynamics (SPH)  $N$ -body code GASOLINE (Wadsley et al. 2004). The system under study comprises a CND, initially in hydrostatic equilibrium, with two embedded BHs initially separated by a distance of the order of the disk’s scale radius. We start by describing the initial conditions setup and then continue with a description of the different physics implemented in the simulations.

### 2.1. Initial Conditions

We used an initial disk model similar to the “smooth set” from Fiacconi et al. (2013), therefore comparable to Escala et al. (2005) and Dotti et al. (2007). The CND structural parameters are set to be commensurate to those of CNDs resulting from galaxy merger simulations (Mayer et al. 2007; Roškar et al. 2015) and observed in gas-rich interacting galaxies (Medling et al. 2014). As a reference, the typical range of molecular gas masses, which is believed to be the dominant gas phase by mass in CNDs, is  $10^8$ – $10^9 M_\odot$  for  $\sim 100$  pc scale CNDs (Izumi et al. 2016). The gaseous disk is represented by  $2 \times 10^5$  SPH (gas) particles, with total mass  $M_d = 10^8 M_\odot$ , and follows a Mestel (1963) surface density profile with scale radius  $R_d = 100$  pc and maximum radius  $\sim 150$  pc, with a Gaussian vertical structure of scale height  $z_d = hR_d$ , where  $h = 0.05$  is the aspect ratio. The initial temperature of all the gas particles is set to  $T_0 = 2 \times 10^4$  K and the initial equilibrium assumes a polytropic equation of state with index  $\gamma = 1.4$ . The disk is embedded in a concentric Plummer (1911) spheroid that models the inner part of a galactic bulge, represented by  $10^6$  star particles, with total mass  $M_P = 5 \times 10^8 M_\odot$ , scale radius  $r_P = 50$  pc, and maximum radius  $\sim 500$  pc.

The primary BH, of initial mass  $M_{01} = 10^7 M_\odot$ , is placed at the center of the CND, then the system is left to relax towards equilibrium for 10 Myr. The secondary BH, of initial mass  $M_{02} = M_{01}/20$ , is positioned at 80 pc

separation from the disk’s longitudinal axis and slightly displaced from the midplane at a height of about 3 pc. Its initial radial and tangential velocity components are both equal to approximately  $110 \text{ km s}^{-1}$ , corresponding to a total initial speed about the same as the local circular velocity and enforcing an eccentric orbit (eccentricity greater than 0.5). The choice of such an eccentric orbit is motivated by the results of pc-resolution galaxy merger simulations modeling the formation of the CND (Mayer et al. 2007; Roškar et al. 2015). The force resolution is determined by a gravitational softening of 0.5 pc, common to all particles.

## 2.2. Radiative cooling, star formation, and supernova feedback

All runs use the same initial configuration of a CND with two embedded BHs, but we greatly vary the sub-grid physics implemented in each simulation (see Table 1). In all simulations, we solve the internal energy equation using the standard density-energy formulation of SPH with a standard Monaghan artificial viscosity (Wadsley et al. 2004) and set the adiabatic index to  $\gamma = 1.4$ , unless otherwise specified (a lower  $\gamma$  leads to a softer equation of state and hence to less compressional/shock heating, which can be an important effect in self-gravitating disks; see e.g. Mayer et al. 2004). We include Compton and non-equilibrium atomic cooling rates for a gas composed of only hydrogen and helium (Wadsley et al. 2004). While metal-line cooling may play a role in the metal-enriched regions at the center of galaxy merger remnants, where CNDs arise, various heating sources would matter as well in the high-star-formation environment expected in such CNDs. These include radiative and photoionization feedback from stellar radiation of OB stars, radiation pressure, photoelectric heating of dust grains and so on, which, in combination with high absorption and scattering opacities in the dense gas, would lead to mean gas temperatures in excess of 1000 K outside cold star-forming clumps, as shown in Roškar et al. (2015), where many of these effects are included. Since our focus here is to explore a relatively large set of simulations, we prefer to limit the numerical complexity of individual simulations and rely on the fact that, with the high resolution of our simulations (below pc scales) and high mean densities of CNDs, the formation of a multiphase medium with dense clumps will occur even with just cooling via recombination and radiative transitions of hydrogen and helium.

Star formation is implemented according to Stinson et al. (2006), therefore using a Schmidt law (Schmidt 1959, 1963) and stochastic conversion of gas particles into star particles for the coldest and densest regions of gas. The star formation efficiency

parameter is set to 0.05 in the Schmidt law. The gas minimum-density threshold for star formation is  $\rho_{\text{thresh}} = 10^5 \text{ amu cm}^{-3}$  and the maximum-temperature threshold,  $T_{\text{SF}}$ , is determined empirically by constraining the global star formation efficiency, i.e. by requiring that the total mass of stars formed over the initial total mass of gas should not exceed 10% over the typical duration of the simulations, consistent with the observations of starbursts in merging galaxies (Medling et al. 2014; note that starburst systems appear to have a slightly higher star formation efficiency relative to normal star-forming galaxies where this is closer to a few percent on a range of scales). The latter is the star formation efficiency resulting when the self-regulating effect of feedback is also included (hence in runs without feedback this can be exceeded significantly over long timescales - see Table 2). This sets  $T_{\text{SF}} = 4500 \text{ K}$ . While this may seem a large temperature threshold, we recall that we do not use metal-line cooling, hence gas will not cool below a few thousand K.

SN feedback also follows the blastwave recipe from Stinson et al. (2006), based on the analytical model described in McKee & Ostriker (1977). Blastwave shocks are assumed to convert kinetic energy into thermal energy on scales smaller than the simulation resolution, therefore only their thermal effects on the surrounding gas are accounted for. Each SN deposits a net energy of  $\epsilon_{\text{SN}} \times 10^{51} \text{ erg}$  in the form of heat into the neighboring gas particles (within the SPH kernel), and we adopt a feedback efficiency  $\epsilon_{\text{SN}} = 0.8$  (Guedes et al. 2011a). Cooling is suppressed for particles within a radius computed via the blastwave solution of McKee & Ostriker (1977) and for a timescale of a few 10 Myr, which corresponds to the snowplow phase. This “delayed cooling” recipe, which has been shown to reproduce realistic galaxies across a wide range of masses (e.g. Guedes et al. 2011b; Shen et al. 2014; Sokołowska et al. 2016b,a), applies only to SNe Type II, while for SNe Type I only a simple thermal energy dump within the local SPH kernel is applied. With the conservative parameters adopted in this paper, our sub-grid model can overestimate the star formation efficiency in the dense ISM of galaxies at high redshift (Sokołowska et al. 2016b), but this is not a concern here as our models are supposed to represent low-redshift CNDs.

## 2.3. Accretion onto black holes

An analytical expression that estimates the hydrodynamic spherical accretion rate onto a non-rotating BH for an infinite mass distribution of gas with no angular momentum is obtained in Bondi (1952) (see also Hoyle & Lyttleton 1939; Bondi & Hoyle 1944). The accretion rate  $\dot{M}_{\text{Bondi}}$  onto a BH of mass  $M_{\text{BH}}$  is expressed in terms of the ambient density  $\rho_{\text{gas}}$ , the sound speed

$c_s$  and the relative velocity  $V$  of the gas with respect to the massive object via the Bondi–Hoyle–Lyttleton (hereafter Bondi) equation:

$$\dot{M}_{\text{Bondi}} = \lambda 4\pi G^2 M_{\text{BH}}^2 \frac{\rho_{\text{gas}}}{(V^2 + c_s^2)^{3/2}}, \quad (1)$$

where  $\lambda$  is a constant of order unity determined by the equation of state of the gas, and  $G$  is the gravitational constant. Our numerical implementation of accretion (Bellovary et al. 2010) considers the parameters in Equation (1) computed within one smoothing length from the BH. In this case,  $\lambda$  is simply a normalization factor, which is intended to reproduce phenomenologically the accretion rates estimated from observational evidence. While in the past large values of  $\lambda$  have been used to compensate for the lack of dense gas due to the low resolution of simulations (e.g. Di Matteo et al. 2005), the sub-pc resolution of our simulations allows us to set  $\lambda = 1$  (see also Callegari et al. 2009; Tamburello et al. 2016). The accretion rate is capped at the Eddington limit. This critical rate,  $\dot{M}_{\text{Edd}}$ , is proportional to the mass of the accreting object,

$$\dot{M}_{\text{Edd}} = \frac{4\pi G m_p}{\epsilon_r \sigma_T c} M_{\text{BH}} = \frac{1}{\tau_{\text{Salp}}} M_{\text{BH}}, \quad (2)$$

where  $m_p$  is the proton mass,  $\epsilon_r$  is an efficiency parameter between 0 and 1 representing the fraction of the mass-energy budget that is assumed to be radiated away (we use  $\epsilon_r = 0.1$ ),  $\sigma_T$  is the Thomson cross section, and  $c$  is the speed of light in vacuum. The time  $\tau_{\text{Salp}} \approx 4.5$  Myr, known as the Salpeter (1964) time, is a proportionality factor associated to a typical timescale during which the BH feeding at this rate grows by a factor of  $e$ . In the simulations, the gas particles that are eligible to be accreted by the BH will have their mass gradually transferred to the compact object according to Equation (1) whenever  $\dot{M}_{\text{Bondi}} < \dot{M}_{\text{Edd}}$ , or based on the rate computed with Equation (2) otherwise.

#### 2.4. AGN feedback

The emitted radiative power in the process of BH accretion (AGN feedback) is assumed to be proportional to the rate at which mass is accreted. A fraction  $\epsilon_{\text{fb}}$  of it heats up the gas around the BH (Bellovary et al. 2010). Therefore, the energy  $E_{\text{fb}}$  released via AGN feedback that thermally couples to the gas per time step  $dt$  can be expressed as

$$E_{\text{fb}} = \epsilon_{\text{fb}} \epsilon_r \dot{M}_{\text{gas}} c^2 dt. \quad (3)$$

In the simulations including AGN feedback, part of the energy released by the accreting BH is thermally coupled to the neighboring gas particles according to Equation (3), with  $\epsilon_{\text{fb}} = 0.05$  (Di Matteo et al. 2005;

Springel et al. 2005; Johansson et al. 2009; Bonoli et al. 2016).

### 3. RESULTS

#### 3.1. Orbital decay

We have performed a small set of adiabatic<sup>1</sup> runs followed by the main set of simulations with radiative cooling, star formation and varying feedback processes. Table 1 describes the simulations used for the analysis appearing in this paper and defines the labels used throughout the text. The BH separation over time is plotted for some representative cases in Figure 1. In the adiabatic runs, the decay occurs on a shorter timescale than that of all the other simulations (except in one case, see below run CSFSN). These runs, whose general behaviour is already well documented in the literature, serve as a benchmark for the subsequent simulations. The main common feature of all the runs that include radiative cooling, irrespective of the presence of other processes, is disk fragmentation. The latter produces a highly inhomogeneous, multiphase background, affecting the orbital decay by making it stochastic, and often leading to a much longer decay timescale (Figure 1). This general outcome confirms the findings of Fiacconi et al. (2013), who used a simple constant-cooling model and had no star formation, BH accretion, nor feedback processes implemented. The degree of clumpiness, however, depends on the specific physics implemented in the runs. In addition, the different physical processes, in particular SN and AGN feedback, affect the CND structure in other ways that can have a subsequent impact on the orbital decay, both globally and locally.

In the remainder we will describe in detail the results of the various runs, implementing an increasingly higher number of physical processes via sub-grid models, starting with a summary of the outcome in the idealized, but highly instructive case of adiabatic runs.

##### 3.1.1. Adiabatic runs

In the adiabatic runs, the disk remains smooth on all scales, and only the perturbation of the secondary BH as it travels through the disk is noticeable (Figure 2).

The BH separation falls to pc-scale in less than 20 Myr (see Figure 1, top panel). During this time, the secondary BH acts as a perturber to the gas, inducing the formation of a wake, and its orbit steadily circularizes (see Figure 2), i.e. the apsides approach the same value, which is why the separation decreases at constant angular momentum (see Figure 1, top panel; and Figure A1).

<sup>1</sup> In this paper, by “adiabatic” we mean runs in which we solve the internal energy equation with cooling switched off.



In this phase the orbital decay is well described by dynamical friction (Chandrasekhar 1943; Ostriker 1999).

**Table 1.** Simulations overview.

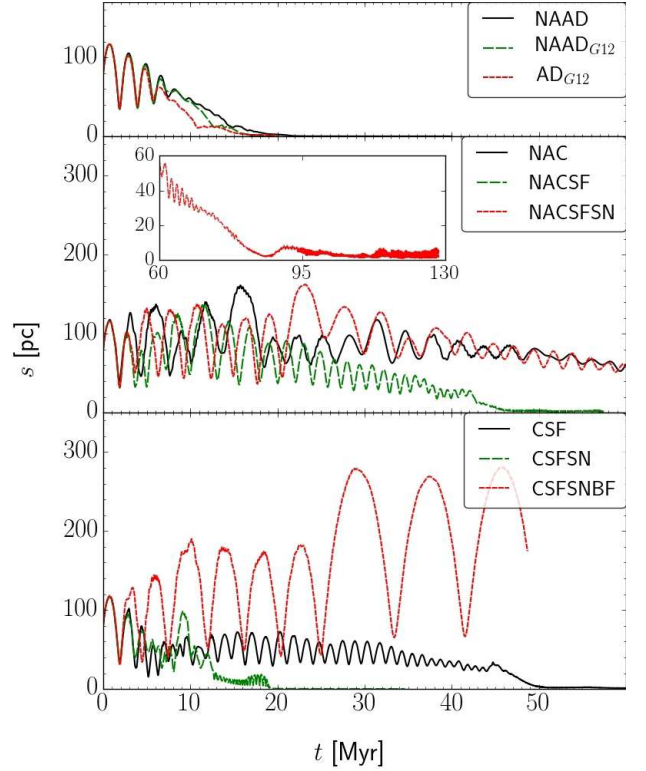
Run label <sup>a</sup>	Accretion	Cooling	Stars	SN	$\epsilon_{\text{fb}}$
NAAD	No	No	No	No	N.a.
NAAD <sub>G12</sub> <sup>b</sup>	No	No	No	No	N.a.
AD <sub>G12</sub> <sup>b</sup>	Yes	No	No	No	0
NAC	No	Yes	No	No	N.a.
NACSF	No	Yes	Yes	No	N.a.
NACSFSN	No	Yes	Yes	Yes	N.a.
C	Yes	Yes	No	No	0
CSF	Yes	Yes	Yes	No	0
CSFSN	Yes	Yes	Yes	Yes	0
CBF	Yes	Yes	No	No	0.05
CSFSNBF	Yes	Yes	Yes	Yes	0.05

<sup>a</sup> NA stands for *no BH accretion*, AD for the *adiabatic* runs, C for *gas cooling*, SF for *star formation*, SN for *SN feedback*, and BF for *feedback* due to accretion onto BHs.

<sup>b</sup> G12 stands for adiabatic/polytropic index 1.2 instead of 1.4.

When gas accretion onto the BHs is considered, circularization happens in a slightly shorter time interval, driving the separation to less than 20 pc faster, although it does not change appreciably the time required to reach (sub-)pc scales, since this is dominated by a second decay phase.

After circularization occurs, indeed, dynamical friction by the background CND, as described in the conventional local-“wake” approach of Chandrasekhar (1943) and Ostriker (1999), becomes weak as the secondary BH has negligible velocity relative to the gaseous background (but not relative to the stellar bulge, which would still exert a drag). Instead, another regime ensues in which spiral density waves are triggered by the secondary BH and back-react by draining its orbital energy and angular momentum. This migration regime is analogous to Type I and Type III migration in planet formation and leads to a faster orbital decay compared to the dynamical friction phase (Mayer 2013). This is characterized by fast extraction of angular momentum from the BH. The negative torques experienced by it have been shown to be associated with two mechanisms, namely the excitation of density waves in the disk near the Lindblad resonances, and the torques from the gas co-orbiting with the BH. The latter co-orbital torques are stronger and cause fast inward migration, as highlighted by the fast steepening of the angular momentum evolution curve in Figure A1 (see Mayer 2013 for torque analysis in a similar case). The short timescale associated with this second stage of the migration process

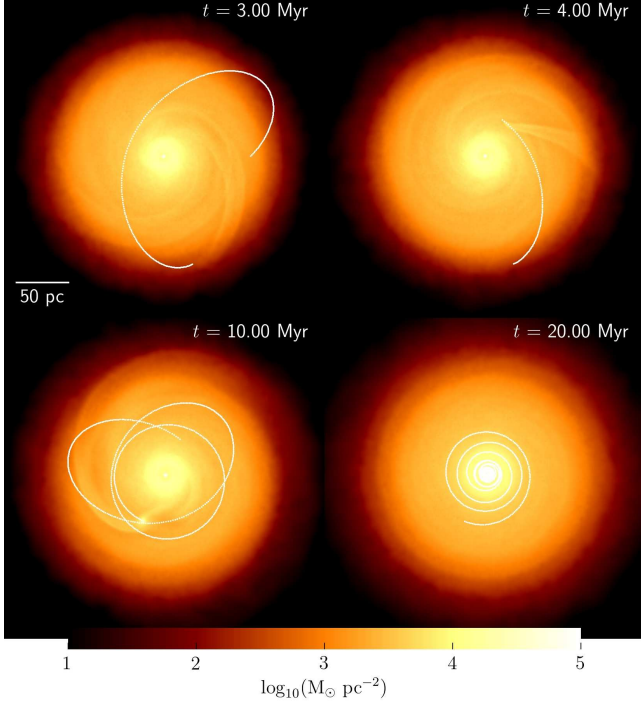


**Figure 1.** Distance between the BHs for different runs (refer to Table 1). The shortest timescales for the orbit of the secondary BH to fall under the one-pc scale separation from the central BH are observed in the adiabatic runs (top panel). Almost all the other runs that include gas cooling lead to prolonged, if present, orbital decay. The exception is run CSFSN (bottom panel), in which an encounter between the secondary BH and an overdense region of gas caused a relatively short decay time. Run CSFSNBF was stopped before 60 Myr due to computational limitations.

likely explains the absence of an evident gap-opening: the secondary BH simply does not have time to develop a gap in regions where the conditions would allow it, as it traverses those too quickly, as thoroughly studied and discussed in Malik et al. (2015) for massive planets and brown dwarfs in protostellar disks.

### 3.1.2. Effects of cooling

We now proceed to study the outcome of the radiative-cooling runs. Cooling, even in absence of metals, is fast at the high densities of the CND, in particular it occurs on a timescale faster than the local orbital time (which is of order of 1 Myr at the half-mass radius of the disk). This leads to rapid fragmentation into clumps, as it is well studied and established in the literature on self-gravitating disks (see e.g. Gammie 2001; Durisen et al. 2007). The transition to a clumpy disk regime is abrupt in absence of any heating and/or star formation, taking place in only 0.3 Myr, namely in a fraction of the time it takes the secondary BH to complete the first orbit (see Figures 3 and 4). At this stage



**Figure 2.** Gas density face-on projection and secondary BH’s trajectory (white lines) for the adiabatic run NAAD (see Table 1) at different times. A scale bar on the top left panel shows the size in pc. The colour code represents the disk density integrated along the  $z$ -axis (perpendicular to the disk plane), which rotates counter-clockwise in this perspective. Trajectories are shown for time intervals between snapshots. On the top left panel, we see a leading spiral wake caused by the BH near an apocenter as a perturbation to the disk density. On the top right panel, a trailing wake is captured near a pericenter. The net effect of the wakes over a few orbits is to circularize the orbit. The lower left panel shows a resonant interaction between the BH and the gas as the orbit becomes circular. After circularizing, the gas extracts its angular momentum and the separation shrinks in a relatively short timescale, as clearly captured on the bottom right panel. The interplay between perturber and background is analogous to results found in the literature for planet migration.

the disk enters a gravo-turbulent regime, in which collisions and tidal interactions between cold clumps occur as well as torquing by strong spiral density waves (Figure 3). Clumps typically gravitationally scatter each other but can also merge into heavier clumps, which can cause even stronger perturbations on the secondary BH. Ejection of the secondary BH from the disk plane can result, and indeed almost always occurs in our simulations (Figures 4 and 5). The overall effect is to induce a stochastic behaviour of the orbital separation as opposed to the smooth behaviour in the non-fragmenting, adiabatic case (Figure 1).

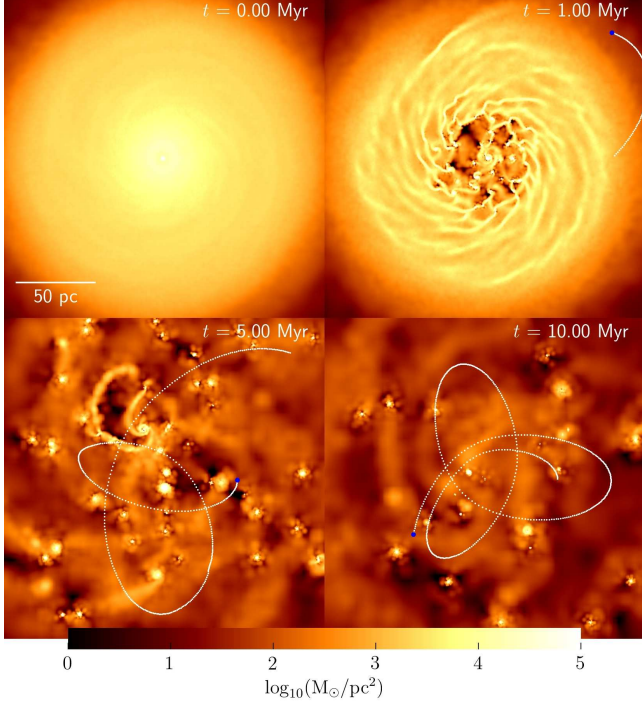
The diffuse background gas is also affected by torques, which produce a global inward flux of mass and make the CND more centrally concentrated. This reflects the nature of self-gravity, which can be described by an effec-

tive viscosity acting throughout the disk even in absence of fragmentation (Lin & Pringle 1987).

In run NAC (see Table 1), within 10 Myr, the inner 25-pc region goes from enclosing less than 20% of the total mass of the disk and being smooth to having more than 50% of the total disk mass in 3 clumps, one of about  $3.19 \times 10^7 M_\odot$  enclosing the primary BH, another one of about  $1.88 \times 10^7 M_\odot$  (heavier than the primary BH itself) at a distance of about 8 pc center to center (each have a radius larger than 1.5 pc), and the third one of about  $6.7 \times 10^6 M_\odot$ , still much heavier than the secondary BH (during this time, the secondary BH is always outside this inner region; see Figure 1).

Figure 6 shows the distribution of clump masses at different times (0.5–5 Myr) normalized to the mass of the secondary BH, which is almost constant during this time range in the runs shown in the figure (NAC, CSFSN, and CSFSNBF; see Table 1). In run NAC, especially after 3–5 Myr, there are more than 10 clumps with masses from commensurate to ten times more massive than the secondary BH. As discussed in Fiacconi et al. (2013) and Roškar et al. (2015), clumps with masses comparable to or larger than the mass of the secondary BH can have a dramatic effect on the latter, with direct hits or even repeated softer perturbations. At the same time, Figure 6 shows that most of the clumps have masses about the same or less than the secondary BH, namely close to or below  $10^6 M_\odot$ , in the range of typical Giant Molecular Clouds (GMCs).

As previously mentioned, the secondary BH is ejected outside the plane of the disk, above the nominal disk scale height. Once that happens, the secondary BH finds itself in a lower density background, which decreases the strength of both dynamical friction and disk torques considerably (see Figure 5). As a result, in run NAC, the separation between the two BHs is still above 50 pc at 60 Myr, in stark contrast to the results of the adiabatic simulations described in the previous section. The redistribution of disk mass due to the gas inflow can in principle have an important effect on the orbital decay of the secondary BH even in absence of an ejection. Indeed, as shown in Figure 7, as the CND becomes more centrally concentrated, the radial profile of the circular velocity of the disk (defined as  $v_{\text{circ}} = \sqrt{M(<r)}/r$ , where  $M(<r)$  is the gas mass inside a radius  $r$ ) becomes much steeper in the center over a few disk orbits inside of 50 pc. The secondary BH will inevitably increase its orbital velocity proportionally to the local circular velocity. Therefore, since dynamical friction is inversely proportional to  $V_{\text{BH}}^2$  in the transonic and supersonic regimes relevant here (see e.g. Mayer 2013), the drag should be significantly reduced. The same effect has been invoked to explain the stalling of sub-pc separation binaries in smooth CNDs resulting from galaxy



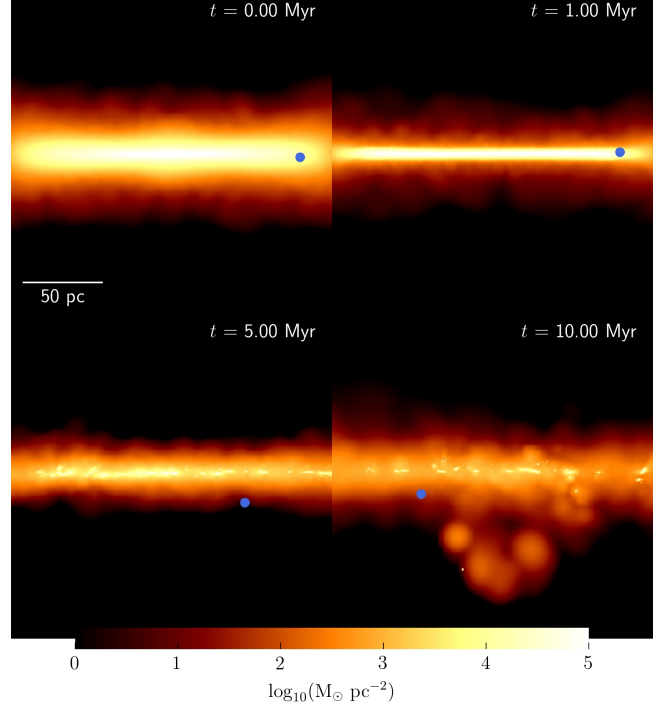
**Figure 3.** Gas density face-on projection and secondary BH's trajectory for run C (see Table 1), similar to Figure 2. Notice the change of the scale and the surface density limits in contrast with that figure. The left and right upper panels show respectively the initial state of the gas and the fragmentation happening in the central regions of the disk. From the bottom left panel we notice that the fragmentation timescale is fairly smaller than the timescale for the secondary BH to complete one orbit. As time elapses, many gaseous clumps are formed and interact with each other, scattering and combining to form fewer larger clumps. The environmental differences along the BH path in comparison to the adiabatic runs are enormous. This leads to a completely different interplay between the gas disk and the BH, as the latter moves through a more contrasted density background.

mergers (Mayer et al. 2008). Figure 7, though, suggests that this central mass accumulation effect is present to some degree in all runs with cooling, irrespective of other processes modeled in the simulation. In the next section we will assess its quantitative effect by comparing the orbital decay in the NAC run with that in the runs with star formation.

Next, we performed a set of different simulations featuring combinations of presence and absence of BH accretion, star formation, and SN feedback. We consider now two representative runs (see Table 1): first CSF (Section 3.1.3) and then CSFSN (Section 3.1.4).

### 3.1.3. Effects of star formation

GMCs are the sites of star formation. Even before feedback processes are considered, the process of conversion of dissipative gas into a dissipationless stellar component should affect disk stability and fragmentation over time, in particular how tightly bound the clumps can get. This can have important effects on the or-

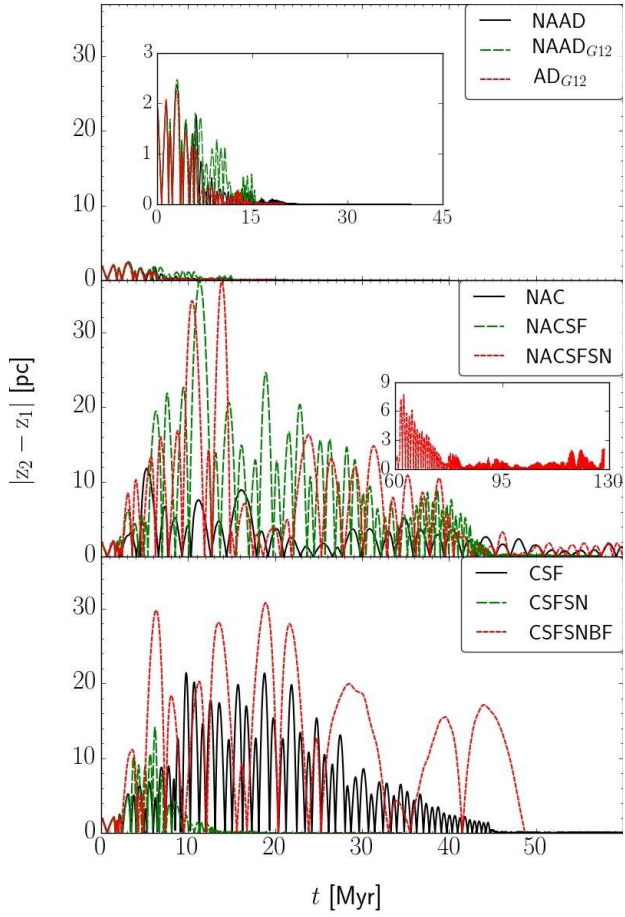


**Figure 4.** Gas density edge-on projection and secondary BH's position for run C (see Table 1). The first panel is the initial state of the disk and is also representative of the vertical structure of the disk in the adiabatic runs for all times. The blue dot indicates the secondary BH position. As the gas cools down, it becomes denser as it offers less pressure support against gravity, finally leading to formation of clumps that are scattered around. On the bottom right panel we can see a returning massive clump that was scattered out far from the disk plane.

bital evolution of the pair of BHs. Indeed, in a scenario of rather efficient star formation, gaseous overdensities form stars rather than continuing to collapse. One expects that, as star formation proceeds, clumps will be less dense and more easily disrupted, reducing the effect of stochastic torques and off-plane scattering. In this section we study the stabilizing effect of star formation in two runs, one without and one with gas accretion onto the two BHs, named, respectively, NACSF and CSF (see Table 1).

In the simulation NACSF, within 10 Myr the amount of newly formed stars and gas in the average clump are both of the order of a few  $10^6 M_\odot$ . Most of the gas has been turned into stars and the CND itself has become a predominantly stellar disk of size similar to the original gaseous one. The rapid conversion of gas into stars not only has suppressed clump formation but also has stabilized the disk globally, leading to a milder evolution of the CND density distribution (Figure 8). Yet the difference in the circular velocity profile at 50 pc is only about 10% (Figure 7), which suggests that the central mass accumulation effect cannot explain the large difference in the orbital decay time between runs NAC and NACSF





**Figure 5.** Vertical separation between the BHs for different runs (refer to Table 1). The plots indicate when ejections out of the disk plane occur. In the top panel (adiabatic runs), the secondary BH never exceeds the 5-pc (initial  $z_d$ ) range from the mid plane.

(or CSF) seen in Figure 1. Therefore, we believe that it is the global disk stabilization by star formation which reduces the amplitude of stochastic torques and brings the orbital decay back to a more regular path similar to the adiabatic case.

Runs with star formation but no feedback are, of course, idealized. In particular, the star formation efficiency is maximized in the absence of feedback. Indeed, as shown by Table 2, towards the end of the simulations most of the disk gas has been turned into stars, while observations of CNDs show that the relative weight of gas and stars in the mass budget is at most comparable (Medling et al. 2014).

Gradual circularization of the BH orbit in the runs NACSF and CSF eventually brings the orbital decay back to the disk-torques driven regime, resulting in a tight pair formation on a  $\sim 50$  Myr time scale (see Figure 1).

Figure 5 shows that, while off-plane scattering of the secondary BH occurs, it does so only in the early stage.

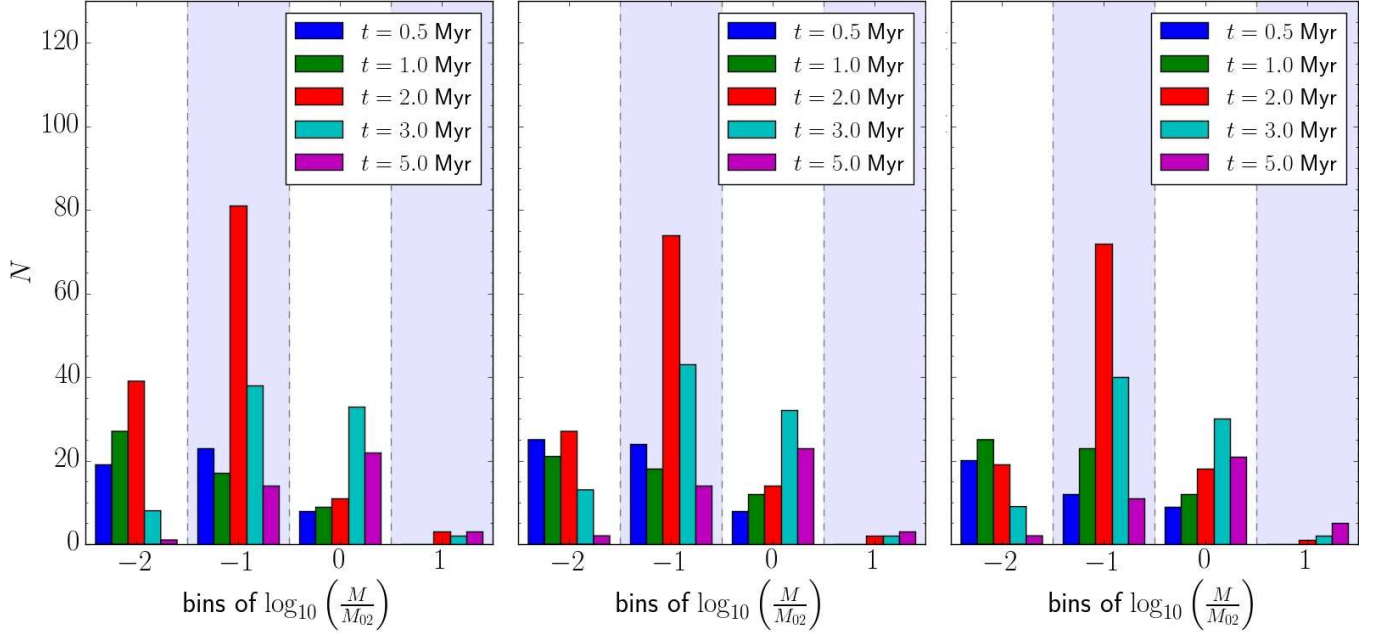
In Figure 1 it is evident that, while initially a delay in the orbital decay occurs relative to the adiabatic runs, the secondary BH ultimately decays on a timescale only a factor of 2 larger relative to the adiabatic runs.

This is consistent with previous studies showing that the orbital decay in smooth disks is negligibly affected by the relative mass budget of gas and stars as long as the CND has the same structural parameters (Dotti et al. 2007). This is clearly applicable to the late stages of the star formation runs, as the CND is substantially smoother. This being said, in run CSF we also observe the capture of a gas clump by the secondary BH after its orbit circularizes, at  $\sim 45$  Myr, which may contribute to accelerating the last phase of the decay (Figure 1) as the secondary acquires a larger effective mass and excites stronger density waves.

#### 3.1.4. Effects of feedback from supernovae

As newly formed massive stars explode as SNII, on timescales of 5–6 Myr and longer, we expect blastwaves to develop, heating and pushing gas away, especially in the direction perpendicular to the disk plane, which offers the path of least resistance as the ambient medium has a much lower density. A considerable amount of gas is lost in outflows (at 45 Myr, 30% of the total remaining amount of gas is found in ejecta in run CSFSN). The effect is evident in Figure A2. Powerful winds develop as the star formation rate, while a factor of 3–4 lower than in the absence of feedback, is still significant compared to the moderate mass of the disk, peaking at  $\sim 2.5 M_\odot \text{ yr}^{-1}$  (Figure 9). Such star formation rates are on the moderate side for CNDs in merger remnants, based on both observations and simulations (Kennicutt 1998; Roškar et al. 2015) while they correspond to the high end of star formation rates found in the CND of low-redshift Seyfert galaxies (Diamond-Stanic & Rieke 2012; Esquej et al. 2014). As inferred by inspecting Figures 1 and 5, the orbital decay is somewhat erratic at the beginning (the disk fragments rapidly and initially behaves as in the ‘cooling and star formation’-only runs until SNII explode) but, after  $\sim 10$  Myr, no ejection occurs. The secondary BH sinks fast to the center, on a timescale of the same order of the adiabatic runs, in one of the runs (CSFSN), while in the other run (NACSFN) it wanders significantly at large distances and then enters a slow-decay phase which, however, does not lead to a successful pairing even after 60 Myr. Note that, while the two runs differ for the presence or not of BH accretion (none of them includes BH feedback), the markedly different result is again caused by the stochastic behaviour in the gravo-turbulent CND. In NACSFN, the inefficient decay is due to the weak drag that the BH suffers once away from the disk midplane, as it finds itself in the low-density envelope produced by SN out-





**Figure 6.** Histogram of the clump mass distribution at different times for runs NAC (left), CSFSN (middle), and CSFSNBF (right). The horizontal axis is shown in terms of the initial mass of the secondary BH, to highlight the potential dynamical influence of a close encounter between the secondary BH and a clump in the disk. As time elapses, smaller clumps merge into larger ones, building up a sensible population of bound objects with mass comparable to or larger than the one of the secondary BH. For reference:  $M_{02} = 5 \times 10^5 M_{\odot}$ . The partition of the horizontal axis is taken from the approximate integer value of the logarithmic function, i.e. masses in the range  $\sim M_{02} \times 10^n / 3.16$  to  $\sim M_{02} \times 10^n \times 3.16$  belong to the bin labelled by  $n$ .

flows at heights above 100 pc (see Figure A2). Instead, the other run suggests that, when the BH is not ejected, the drag in the CND is of comparable strength to the cases without feedback, as similar decay timescales are obtained, despite how the CND density distribution and temperature is affected.

Furthermore, here again we observe an eventual tendency of orbital circularization for both the clumps and the BH trajectories (Figure 1). This favours the occurrence of co-orbital encounters between them at late stages of the simulation (after 35 Myr). At this stage, several of those encounters eventually led to the merging between the gas clouds and the BH involved, likely explaining the sudden fast decay in the late stage of the CSFSN run, which includes BH accretion.

### 3.1.5. Effects of AGN feedback

When AGN feedback is implemented, the overall disk morphology is not too different from the run without AGN feedback (Figure A2). However, as it will be explained below, local effects of BH feedback on the interstellar medium play an important role on the orbital decay of the secondary BH.

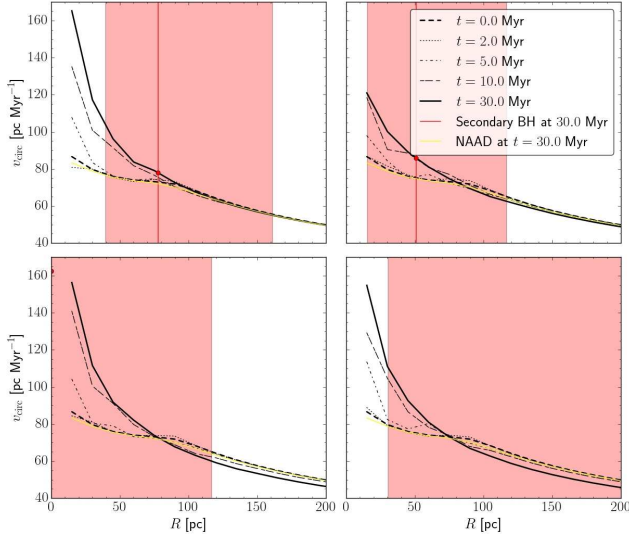
Gas outflows from the CND become slightly more prominent than with SNII blastwave feedback alone, developing strong bubbles of very low density and extremely hot gas after several Myr (Figure 10), which

reach out to larger distances from the disk midplane compared to the runs with only SN feedback.

After 10 Myr, extensive, very hot bubbles reach out to more than kpc scales, creating a low-density diffuse warm/hot envelope (Figure A2) in which dynamical friction onto the secondary BH would be very weak.

The outflow, though, is weakly loaded in mass, hence does not have a considerable impact on the mean properties of the CND (see Table 2). In both runs CSFSN and CSFSNBF, the total baryonic mass ejected from the disk (defined as a box of side length 300 pc and 20 pc height centered at the primary BH) at  $t = 20$  Myr is, respectively,  $\sim 14\%$  and  $\sim 15\%$ . This is consistent with the fact that the circular velocity profiles (see Figure 7) suggest that the gravitational potential wells of the disks themselves are similar between those two runs up to  $t = 30$  Myr. Similarly to the runs with SN feedback and no AGN feedback, star formation is essentially truncated after about 10 Myr (Figure 9), despite the fact that a large reservoir of gas remains (Table 2). The distribution of clump masses during the initial 5 Myr of the simulation, indeed, does not differ much from that of runs without AGN feedback (see Figure 6).

These findings may seem in contrast with the expectation that an extra source of energy feedback would suppress further gas densities and subsequently have an impact on star formation rates. On the other hand, it



**Figure 7.** Circular velocity profiles centered on the primary BH’s position for runs NAC (top-left), CSF (top-right), CSFSN (bottom-left), and CSFSNBF (bottom-right). The red vertical line indicates the radial distance, projected on the disk plane, of the secondary BH from the primary one at  $t = 30$  Myr, whereas the shaded region indicates the range of such distances during the first 30 Myr.

has been reported that the energy output from AGN in a dense disk will have little impact on the disk itself (Roos et al. 2015), especially when compared to the energy that couples to the disk coming from stars that go SNe, which is distributed all over the disk. This does not mean that AGN feedback does not have important consequences; it regulates BH growth, and it introduces a new effect that can alter the orbital decay timescale of the secondary BH, as we are going to explain.

First, AGN feedback has a strong self-regulation action on the mass growth of the secondary BH. Without AGN feedback, the accretion rate is always Eddington-limited as, with a dense gas reservoir, the Bondi rate would often exceed the Eddington rate, while with AGN feedback it leads to Eddington rates of  $\lesssim 0.2$ . For the primary BH, instead, growth remains Eddington-limited also with AGN feedback, suggesting that, in the very dense nucleus of the CND, AGN feedback is not as effective at self-regulating its mass growth in the way it is implemented in our simulations.

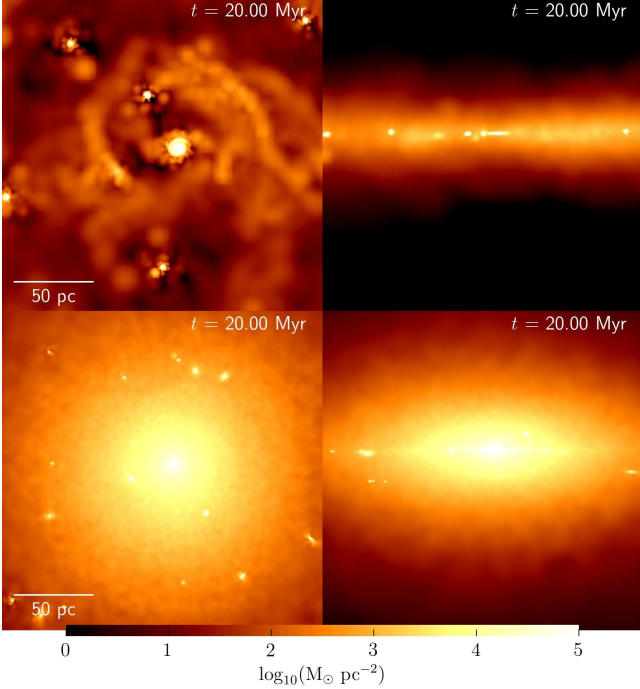
**Table 2.** Baryonic mass at 20 Myr. By “mass in a disk” in this table, we mean the mass within a box centered at the center of mass of the system with 20 pc height and 300 pc side lengths.  $M_g$  and  $M_*$  are calculated at 10 Myr, whereas  $M_d$  is the initial disk mass.

Simulation	$M_g/M_d$	$M_*/M_g$	$M_{BH1}/M_{01}$	$M_{BH2}/M_{02}$
NAC	0.9766	0.0	1.0	1.0
CSF	0.1416	4.4544	1.1081	1.0718
CSFSN	0.7451	0.0774	1.5569	1.3137
CSFSNBF	0.7423	0.0745	1.547	1.0026

Second, AGN feedback causes what we dub the *wake evacuation* effect. For run CBF, which allows to isolate the effects of AGN feedback from SN feedback, it is clearly seen in Figure A3 (top panels). What happens is that, as the secondary BH undergoes a first accretion burst, the bubble produced by AGN feedback primarily evacuates gas behind the secondary BH and trailing with it, which reduces the effect of dynamical friction. The effect plays a role as long as the secondary BH is in the CND. Even after an out-of-plane scattering occurs, it is still noticeable when the secondary BH crosses again the CND along its orbit. The wake evacuation effect explains why, even before any ejection occurs, the orbital decay is really inefficient in runs with AGN feedback (see Figure 1 before 10 Myr). Ejections out of the CND plane occur in all our runs with AGN feedback, and in some are the most dramatic in our sample, sending the BH to more than 280 pc from the CND center (e.g. see run CSFSNBF in Figure 1). Wake evacuation effects and ejections combine uniquely in AGN feedback runs to render the orbital decay more inefficient than in all the other runs. Indeed as BHs, due to the slow decay induced by the wake evacuation effect, are ejected when they are still very far from the center of the disk, inevitably their orbital decay is maximally suppressed because they need to lose a larger fraction of their initial orbital energy and orbital angular momentum (Figure A2). Indeed, the secondary BH (in the late stages of run CSFSNBF) keeps orbiting the center at distances between 60 and 280 pc (until we stopped the simulation). In the next section, where we estimate the dynamical friction effect of a large bulge extending out to such large distances, we will discuss the possible timescale for the secondary BH to return to the plane and form a bound BH pair at the center.

### 3.2. Estimate of the orbital decay timescale

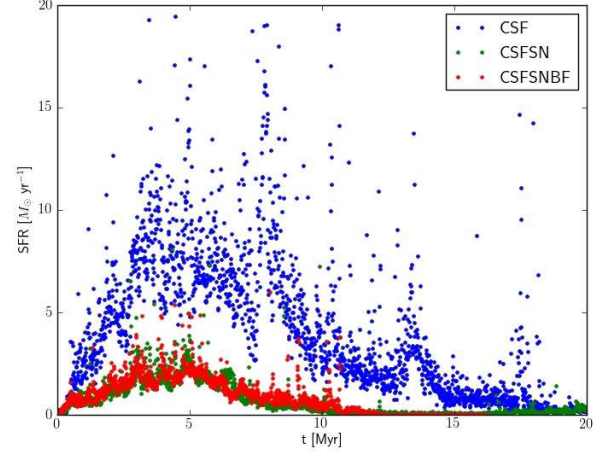
Encounters of the secondary BH with clumps in eccentric orbits can eject the BH out of the disk. We have seen several examples of this phenomenon in our sim-



**Figure 8.** Baryonic density maps of runs NAC and NACSF. In the top row we show the face-on (left) and edge-on (right) gas density maps at 20 Myr for run NAC. The bottom row shows the corresponding maps for the stellar density in run NACSF, which is the dominant component of the disk’s baryonic content (in this run, the stellar mass surpasses the gaseous mass by a factor larger than 4). Strong star formation leads to a less contrasted density background along the secondary BH trajectory, as the collapsing massive gas clumps give space to stellar overdensities that are less likely to further collapse into a more closely bound structure. By comparatively loosening up the clumps, they become more prone to disruption via encounters, allowing the matter to be redistributed into a smoother disk. A smoother background remits to the adiabatic runs, in the sense that the dynamics of the secondary BH becomes dominated by its interaction with the disk on a local and global scale, whereas strong scatterings from clump encounters play a minor role.

ulations. Once outside the CND’s midplane, the low-density ambient medium implies very long dynamical-friction timescales. Indeed, our simulations show cases in which the BH pairing appears to stall, in the sense that the separation evolves little or even increases in some cases after tens of Myr. However, if we had included an extended stellar bulge in our simulations, this would have exerted dynamical friction onto the secondary BH even well outside the CND. Note that the stellar (Plummer) bulge already included in our simulation setup only represents a small fraction of the total extent of the bulge of a massive spiral or galaxy merger remnant, its scale radius being more typical of a galactic stellar nucleus (50 pc). This choice was dictated by reasons of computational efficiency.

We can thus estimate what effect would an extended bulge have on the orbital decay timescales down to the



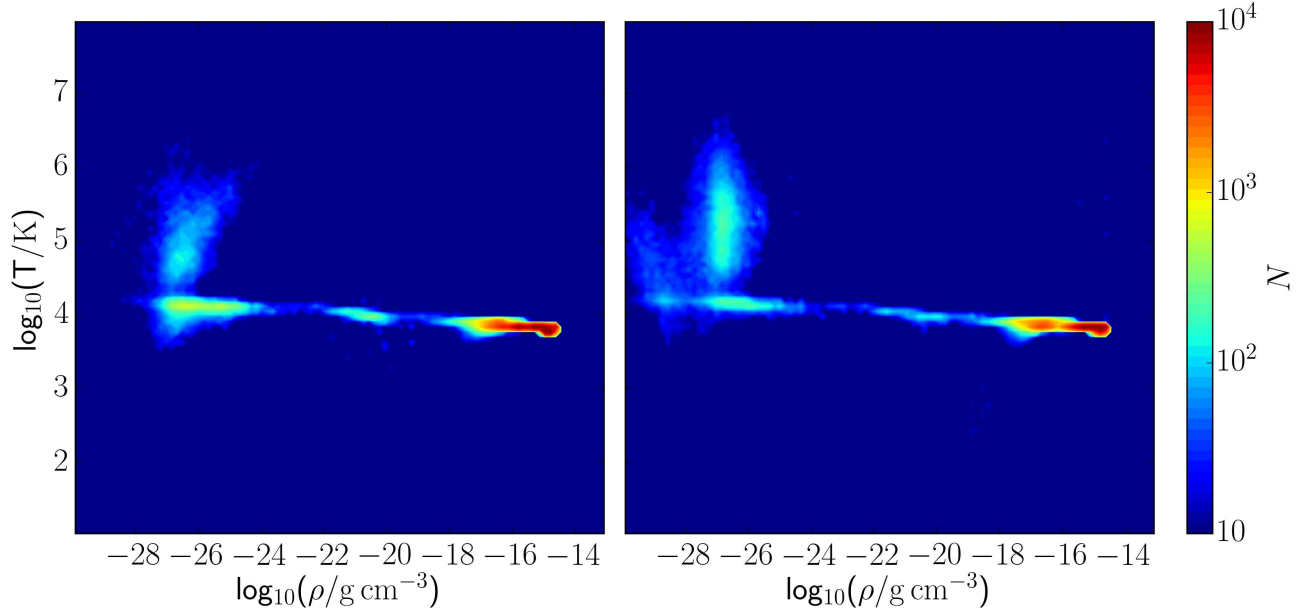
**Figure 9.** Effects of the feedback mechanisms on the star formation rates in the simulations. SN feedback is the key ingredient that suppresses higher star formation rates. The introduction of AGN feedback has a negligible effect in that sense.

formation of a hard BH binary<sup>2</sup>. As we show below, we find an upper limit of  $\sim 3 \times 10^8$  yr for the orbital decay to negligibly small separations, where emission from gravitational waves would take place, if we consider that the decay is driven by dynamical friction from a fictitious [Hernquist \(1990\)](#) bulge. We further have to assume that the BH is moving in a circular orbit and that no subsequent ejections occur. This is a rather long timescale, a fraction of a Gyr, but it is only a factor of 3 longer than the timescale obtained in the merger simulations of [Roškar et al. \(2015\)](#), where extended stellar and dark matter distributions were present by construction, but in which many of the processes studied in this paper, in particular BH accretion and feedback, were not taken into account. The likelihood of the occurrence of further ejections is clearly related to how clumpy the disk will be by the time it recaptures the secondary BH, which depends upon the lifespan of massive clumps and on the timescale in which the nuclear region of a galaxy is accreting gas and replenishing its gas reservoir (or not, if the effect of feedbacks is very strong). Since in the simulations we consider a disk in isolation without gas inflows from the outer regions of the host galaxy, clumps inevitably become more rare after several tens of Myr.

We resort to the Hernquist analytical model in an attempt to consider an extended bulge for our estimate. By taking into account the effects of dynamical friction in a fictitious stellar Hernquist bulge of scale radius  $a$  and total mass  $M_b$ , the corresponding timescale for a

<sup>2</sup> When the BHs inter-separation remains under 5 pc after several orbits.





**Figure 10.** Phase diagrams for runs CSFSN (left) and CSFSNBF (right) at 30 Myr. The color coding shows particle number, so that brighter regions have more particles in the corresponding range of density and temperature. The overall morphology is similar between the two runs but a larger fraction of the gas clearly can reach higher temperatures in the AGN feedback case, as shown by the fact that the diagram is more densely populated above  $10^{5.5}$  K. Other differences are likely related to the intrinsic stochasticity of star formation and SN feedback.

body of mass  $M$  in circular motion inside such a bulge to inspiral from an initial radial distance  $r_i$  to a final one  $r_f$  is given by

$$t_{\text{df}} = \frac{1}{4\xi \ln \Lambda} \sqrt{\frac{a^3}{GM_b}} \frac{M_b}{M} \int_{\chi_f}^{\chi_i} \frac{\chi^{3/2}(\chi+3)}{\chi+1} d\chi, \quad (4)$$

where  $\xi$  is a correction factor of order unity,  $\ln \Lambda$  is the Coulomb logarithm, and  $\chi = r/a$  is the ratio between the radial coordinate of the orbiting mass,  $r$ , and the bulge's scale radius,  $a$ . Under the assumption that  $\chi_i \ll 1$  (i.e.  $r_i \ll a$ ), the integral in Equation (4) can be approximated by  $6\chi^{5/2}/5$ , and we can re-write the equation as

$$t_{\text{df}} = 372 \xi^{-1} \text{ Myr} \left( \frac{\ln \Lambda}{5} \right)^{-1} \left( \frac{M_b}{10^{10} M_\odot} \right)^{1/2} \times \left( \frac{M}{5 \times 10^5 M_\odot} \right)^{-1} \left( \frac{a}{750 \text{ pc}} \right)^{3/2} \left( \frac{r_i/a}{0.4} \right)^{5/2}. \quad (5)$$

For the parameters of the bulge, we consider  $a = 750$  pc and  $M_b = 10^{10} M_\odot$ . For the secondary BH, we take  $M = 5 \times 10^5 M_\odot$  (the initial mass of the BH),  $r_i = 300$  pc, and  $r_f = 0$  pc. The choice of  $r_i$  is motivated by the fact that the secondary BH can be almost this far from the center near apocenter in some of the runs, e.g. those with AGN feedback, in which strong ejections occur. Taking  $\ln \Lambda = 5$  (see [Binney & Tremaine 2008](#))

and  $\xi = 1$ , we find (using Equation 4)

$$t_{\text{df}} \simeq 0.3 \text{ Gyr}. \quad (6)$$

Since in many of our simulations ejections occur to smaller distances than 300 pc, the upper limit on the orbital decay timescale will change accordingly. For instance, a large fraction of runs have secondary BHs ejected to 100–150 pc (see Figure 1), which would yield an expected BH hard binary formation timescale of the order of  $2\text{--}7 \times 10^7$  yr (of the order of what found in [Roškar et al. 2015](#)). However, for the AGN feedback runs the 300 pc limit is representative as ejections occur to comparable distances. In the next section we comment on the implications of such rather long timescales.

#### 4. SUMMARY AND DISCUSSION

The suite of CND simulations with embedded massive BH pairs presented in this paper represents the most comprehensive study so far of all the physical processes that can affect the orbital decay of massive BHs from 100-pc separations to separations of fractions of pc. Looking at the essential picture emerging from the simulations, it is clear that the orbital decay timescale is generally increased relative to the adiabatic runs, but also that it can span a really wide range, from 20 Myr to an estimated maximum of more than 300 Myr (given that no further ejections occur).

*First*, our simulations confirm that, as long as radiative cooling is included, the CND becomes highly

clumpy due to fragmentation, with masses of clumps as large as  $10^7 M_\odot$ , which corresponds to the largest GMC masses, in essentially all the runs despite the inclusion of other processes such as feedback (Figure 6). Clumps and overdense spiral arms introduce a *stochastic orbital decay* regime, leading to perturbations as well as direct hits to the secondary BH, which almost invariably scatter it out of the CND plane. Once outside the CND plane, dynamical friction is considerably reduced, increasing the orbital decay timescale by at least an order of magnitude. This is the dominant effect in all runs, confirming the results of Fiacconi et al. (2013) and Roškar et al. (2015).

*Second*, as the CND becomes unstable, mass flows inward, steepening the CND density and circular velocity profile relative to the adiabatic case. This steepening, however, has relatively little effect on the orbital decay of the secondary BH, compared to, e.g., clumpiness.

*Third*, star formation and feedback processes decrease the gas content of the CND. Feedback leads to outflows that produce a diffuse hot envelope around the CND. However, globally outflows do not have a measurable impact on the orbital decay as most of the mass stays in the CND midplane. Likewise, the orbital decay in a mostly stellar CND is almost identical to that in a gaseous CND as long as densities are comparable, which also confirms previous results (Dotti et al. 2007).

*Fourth*, The warmer medium produced by feedback processes in the CND, though, also concurs to reduce the strength of the drag, especially in the dynamical friction phase. The extreme manifestation of this is the *wake evacuation effect* in AGN feedback runs, in which a hot bubble driven by feedback around the secondary BH lowers drastically the gas density behind the secondary, where the dynamical friction wake should form, stifling orbital decay.

Regarding the first effect, we notice that there is a subset of runs in which clumps can actually have a positive feedback, accelerating the orbital decay time. For example, in run CSFSN a clump scattered the BH towards the center, promoting a shorter decay. We also observed post-orbit-circularization co-orbital encounters between clumps and the BH, much more common at later stages of the simulations (depending on the run, but roughly after 30 Myr). These can also aid decay as the secondary BH accretes mass efficiently from a clump or even merges with it, reaching a higher mass that amplifies the negative torque responsible for its decay. Some gas can remain in a sub-disk around the secondary BH, which can also have a positive effect on the decay as it torques directly onto the sub-disk formed around the primary BH in the last stage of the decay (yet this effect will have to be studied better with higher-resolution simulations delving much deeper in the sub-pc regime of

the decay).

Our simulations do have a number of caveats. To mention a few, the radiative cooling model used in this work is less sophisticated than that employed by Roškar et al. (2015), as it lacks to account for metal lines above  $T = 10^4$  K, molecular and dust-gas cooling, as well as self-shielding at high densities/opacity. However, high-temperature metal cooling is less important in isolated CND simulations as there are no shocks as in a galaxy merger, yet it could affect SN and AGN-driven hot outflows by cooling gas faster out of the bubbles. Nevertheless, since outflows have little effect on the orbital decay, we do not expect this to be a major limitation for our purpose. Self-shielding would reduce cooling at the highest densities, but since we also do not include molecular cooling or cooling via dust-gas collisions, which dominate at the highest densities, the new result should be similar to what we model here. Indeed the masses and characteristic densities of clumps in our simulations are similar to those in Roškar et al. (2015). In the literature there are also newer prescriptions for BH accretion and feedback which take into account angular momentum loss by torques or effective viscosity in accretion disks, which is a more realistic approach relative to Bondi accretion. Some of these recipes have been shown to lead to smaller accretion rates relative to Bondi. However, since the secondary BH accretes already at markedly sub-Eddington rates with AGN feedback, such recipes should have little impact on our results (they would on the growth of the primary BH, though, but this has no relevance for orbital decay). Our systems were considered in isolation, therefore there is room to address the effects of galactic gas replenishing the nuclear disk. And, of course, a future, even larger suite of runs should explore varying initial conditions parameters such as the mass of the disk, the initial mass of the secondary BH or its orbital inclination with respect to the disk. The latter approach was used in some of the previous works appeared in the literature (e.g. Fiacconi et al. 2013; del Valle et al. 2015).

The BH masses considered in this work are within the detection window of eLISA (Amaro-Seoane et al. 2013), hence the orbital decay timescales that we estimate are of high interest. BH tight binary formation timescales ranging from a few times  $10^7$  to a few times  $10^8$  yr are in contrast with the once prevailing idea that orbital decay should be accelerated in gaseous backgrounds as opposed to stellar backgrounds (Mayer et al. 2007; Mayer 2013). A similar delay effect due to a clumpy interstellar medium has been studied recently also at larger galactic scales (Tamburello et al. 2016). However, while in the latter case disk fragmentation into clumps might be important only in the most massive galaxies at high redshift, in the dense nuclear regions studied here this

behaviour appears to be very generic, being now supported by several calculations done with different setups and simulation techniques (Fiacconi et al. 2013; Roškar et al. 2015; del Valle et al. 2015). Here we have enriched the scenario by including new effects induced by feedback processes, most notably wake evacuation, that can delay the decay even in absence of ejections. On the other hand, Khan et al. (2016) have recently shown, using cosmological simulations, that the orbital decay can proceed really fast, in a few Myr, in the dense cores of galaxies at  $z > 2$  when such cores are devoid of gas and 3-body encounters with stars govern the decay process. Putting these two results together, it would appear that, while gas dissipation is needed to create dense cores in galaxies, a fast decay requires at the same time that such gas dissipation stops being important when the BH pair has been delivered to the nucleus after the merger. This is mimicked in an idealized way by our runs with only star formation and no feedback effects in that they maintain a prolonged phase where dissipative gas physics and a clumpy interstellar medium exist. And, indeed, in the latter case the orbital decay is only marginally longer than in the adiabatic, non-dissipative gas cases. Also, it would be shortened further by the concurrent effect of 3-body interactions, that we do not simulate here.

This suggests that there might be at least two regimes to take into account. Namely, BH mergers might be less efficient in gas-rich galaxies, which are typically disk-dominated at all radii, and be more efficient in galaxies that have already built a substantial bulge, in which

the nuclear gas content is expected to be low. It remains to be understood what this implies in terms of the eLISA discovery space, since one can imagine a variety of galaxy hosts, more or less affected by dissipation, which could harbour BHs with masses in the range relevant for gravitational wave detection. Nevertheless, our results show that future studies making predictions for massive BH merger rates must somehow take into account the nature of galaxy hosts and the fact that there might be different regimes of orbital decay depending on the properties of such hosts.

On the other side of the coin, a longer decay time in the 100-pc to pc-scale separation is interesting for observations of tight double AGN. Studies of dual AGN activity in galaxy mergers (Van Wassenhove et al. 2012, Capelo et al., in prep.) have already shown that strong dual AGN activity occurs at relatively small separations ( $\lesssim 1$  kpc). As many of our BH pairs will remain at  $\sim 100$ -pc separations for fractions of a Gyr, this prompts searches for double AGN at small separations in gas-rich, disk-dominated galaxies with small bulges. Yet, as the secondary BH, once ejected, is harboured in the low density, hot envelope around the CND, accretion is suppressed, and so would dual AGN activity. Possible dual AGN activity in this regime will have to be studied in future work.

## 5. ACKNOWLEDGEMENTS

This work is supported by the STARFORM Sinergia Project funded by the Swiss National Science Foundation. RSL thanks Davide Fiacconi for his help. PRC acknowledges support by the Tomalla Foundation.

## REFERENCES

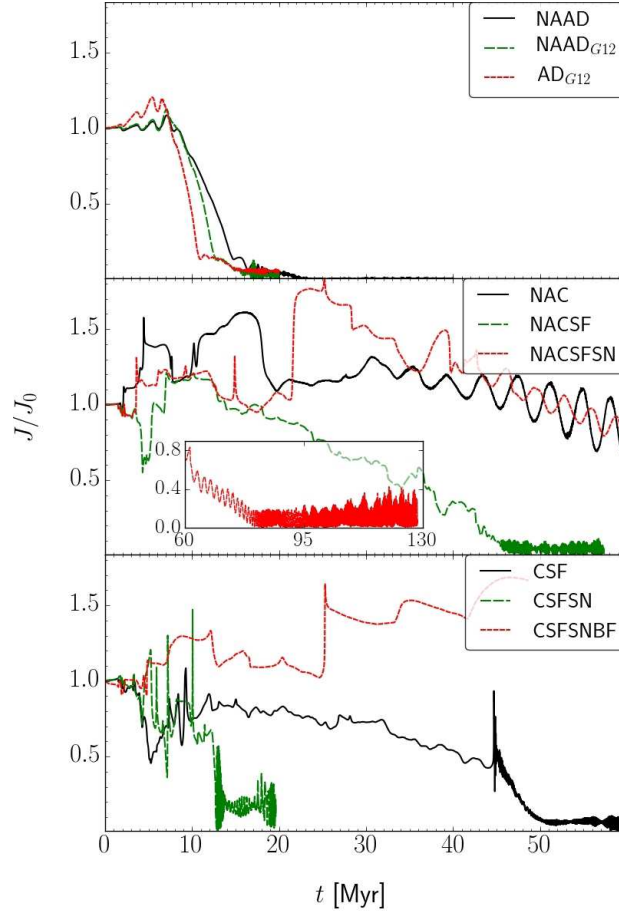
- Agertz, O., Lake, G., Teyssier, R., et al. 2009, *MNRAS*, 392, 294  
Amaro-Seoane, P., Aoudia, S., Babak, S., et al. 2013, *GW Notes*, Vol. 6, p. 4-110, 6, 4  
Begelman, M. C., Blandford, R. D., & Rees, M. J. 1980, *Nature*, 287, 307  
Bellovary, J. M., Governato, F., Quinn, T. R., et al. 2010, *ApJL*, 721, L148  
Binney, J., & Tremaine, S. 2008, *Galactic Dynamics* (Second Edition), 2nd edn., Vol. 1 (Princeton University Press)  
Bondi, H. 1952, *MNRAS*, 112, 195  
Bondi, H., & Hoyle, F. 1944, *MNRAS*, 104, 273  
Bonoli, S., Mayer, L., Kazantzidis, S., et al. 2016, *MNRAS*, 459, 2603  
Callegari, S., Mayer, L., Kazantzidis, S., et al. 2009, *ApJL*, 696, L89  
Capelo, P. R., & Dotti, M. 2016, *MNRAS*, submitted  
Capelo, P. R., Volonteri, M., Dotti, M., et al. 2015, *MNRAS*, 447, 2123  
Chandrasekhar, S. 1943, *ApJ*, 97, 255  
del Valle, L., Escala, A., Maureira-Fredes, C., et al. 2015, *ApJ*, 811, 59  
Di Matteo, T., Springel, V., & Hernquist, L. 2005, *Nature*, 433, 604  
Diamond-Stanic, A. M., & Rieke, G. H. 2012, *ApJ*, 746, 168  
Dotti, M., Colpi, M., Haardt, F., & Mayer, L. 2007, *MNRAS*, 379, 956  
Downes, D., & Solomon, P. M. 1998, *ApJ*, 507, 615  
Durisen, R. H., Boss, A. P., Mayer, L., et al. 2007, *Protostars and Planets V*, 607  
Escala, A., Larson, R. B., Coppi, P. S., & Mardones, D. 2005, *ApJ*, 630, 152  
Esquej, P., Alonso-Herrero, A., González-Martín, O., et al. 2014, *ApJ*, 780, 86  
Fiacconi, D., Mayer, L., Roškar, R., & Colpi, M. 2013, *ApJL*, 777, L14  
Gammie, C. F. 2001, *ApJ*, 553, 174  
Guedes, J., Callegari, S., Madau, P., & Mayer, L. 2011a, *ApJ*, 742, 76  
Guedes, J., Madau, P., Mayer, L., & Callegari, S. 2011b, *ApJ*, 729, 125  
Hernquist, L. 1990, *ApJ*, 356, 359  
Hoyle, F., & Lyttleton, R. A. 1939, *Mathematical Proceedings of the Cambridge Philosophical Society*, 35, 592  
Izumi, T., Kawakatu, N., & Kohno, K. 2016, *ApJ*, 827, 81  
Johansson, P. H., Naab, T., & Burkert, A. 2009, *ApJ*, 690, 802  
Kennicutt, Jr., R. C. 1998, *ARA&A*, 36, 189



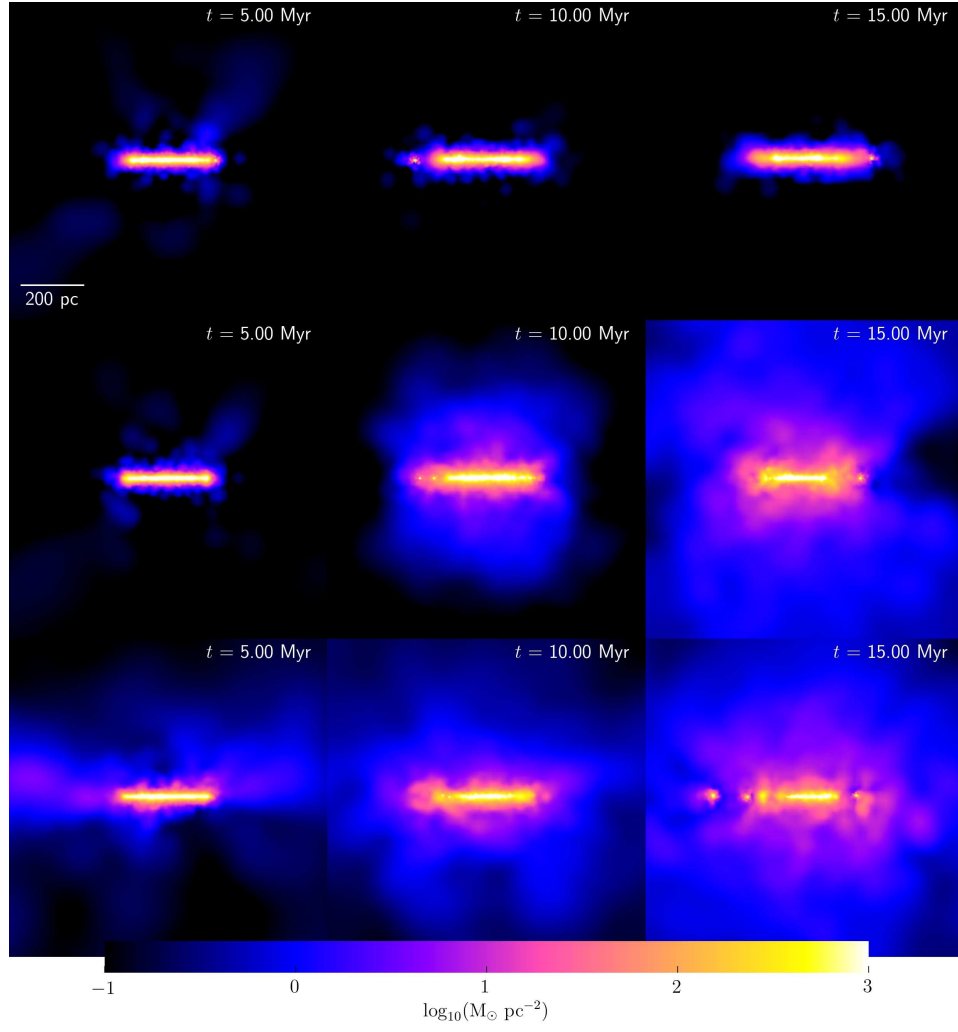
- Khan, F. M., Fiacconi, D., Mayer, L., Berczik, P., & Just, A. 2016, ArXiv e-prints, arXiv:1604.00015
- Klessen, R. S., Glover, S. C. O., Clark, P. C., et al. 2010, in American Institute of Physics Conference Series, Vol. 1294, American Institute of Physics Conference Series, ed. D. J. Whalen, V. Bromm, & N. Yoshida, 28–33
- Lin, D. N. C., & Pringle, J. E. 1987, MNRAS, 225, 607
- Lupi, A., Haardt, F., Dotti, M., & Colpi, M. 2015, MNRAS, 453, 3437
- Magorrian, J., Tremaine, S., Richstone, D., et al. 1998, AJ, 115, 2285
- Malik, M., Meru, F., Mayer, L., & Meyer, M. 2015, ApJ, 802, 56
- Mayer, L. 2013, Classical and Quantum Gravity, 30
- Mayer, L., Kazantzidis, S., & Escala, A. 2008, Mem. S.A.It., 79, 1
- Mayer, L., Kazantzidis, S., Escala, A., & Callegari, S. 2010, Nature, 466, 1082
- Mayer, L., Kazantzidis, S., Madau, P., et al. 2007, Science, 316, 1874
- Mayer, L., Quinn, T., Wadsley, J., & Stadel, J. 2004, ApJ, 609, 1045
- McKee, C. F., & Ostriker, J. P. 1977, ApJ, 218, 148
- Medling, A. M., U, V., Guedes, J., et al. 2014, ApJ, 784, 70
- Mestel, L. 1963, MNRAS, 126, 553
- Ostriker, E. 1999, ApJ, 513, 252
- Plummer, H. C. 1911, MNRAS, 71, 460
- Richstone, D., Ajhar, E. A., Bender, R., et al. 1998, Nature, 395, A14
- Roos, O., Juneau, S., Bournaud, F., & Gabor, J. M. 2015, ApJ, 800, 19
- Roškar, R., Fiacconi, D., Mayer, L., et al. 2015, MNRAS, 449, 494
- Salpeter, E. E. 1964, ApJ, 140, 796
- Schmidt, M. 1959, ApJ, 129, 243
- Schmidt, M. 1963, Nature, 197, 1040
- Shen, S., Madau, P., Conroy, C., Governato, F., & Mayer, L. 2014, ApJ, 792, 99
- Sokolowska, A., Capelo, P. R., Fall, S. M., et al. 2016a, ApJ, submitted
- Sokolowska, A., Mayer, L., Babul, A., Madau, P., & Shen, S. 2016b, ApJ, 819, 21
- Springel, V., Di Matteo, T., & Hernquist, L. 2005, MNRAS, 361, 776
- Stinson, G., Seth, A., Katz, N., et al. 2006, MNRAS, 373, 1074
- Tamburello, V., Capelo, P. R., Mayer, L., Bellovary, J. M., & Wadsley, J. 2016, ArXiv e-prints, arXiv:1603.00021
- Tasker, E. J., & Tan, J. C. 2009, ApJ, 700, 358
- Van Wassenhove, S., Capelo, P. R., Volonteri, M., et al. 2014, MNRAS, 439, 474
- Van Wassenhove, S., Volonteri, M., Mayer, L., et al. 2012, ApJL, 748, L7
- Wadsley, J. W., Stadel, J., & Quinn, T. 2004, New Astronomy, 9, 137
- White, S. D. M., & Rees, M. J. 1978, MNRAS, 183, 341

## APPENDIX

### A. ADDITIONAL FIGURES

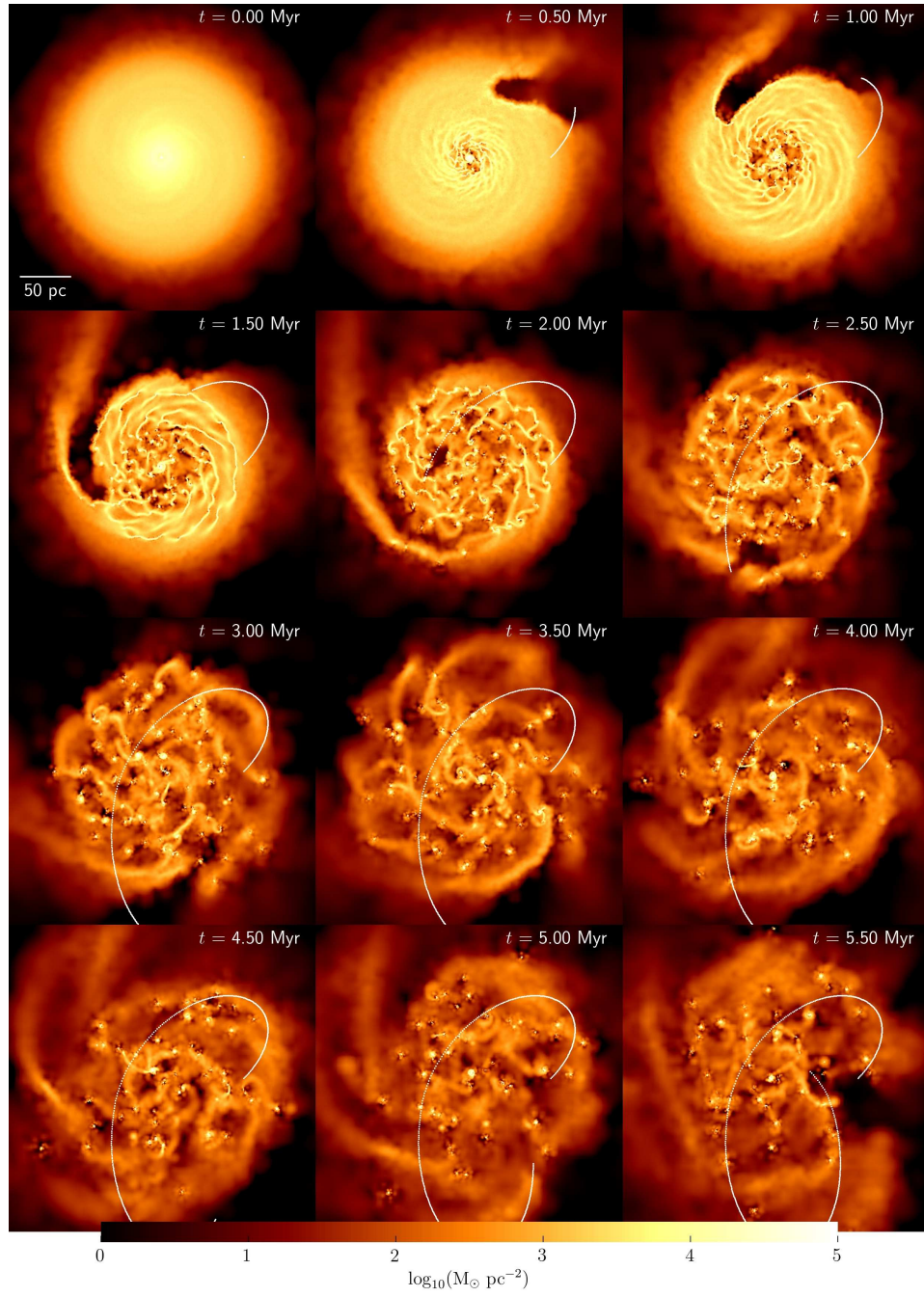


**Figure A1.** Angular momentum of the secondary BH with respect to the center of mass of the entire system (all particles) for different runs (refer to Table 1). We have to bear in mind that ultimately we want to address the problem of the angular momentum of the system composed by the two BHs (in whichever external forces field they may lie on). This is *not* what is shown in this plot. Since the primary BH is very massive and lies close to the center of the whole system, showing the right quantity would only introduce the noise caused by the primary BH's motion. Here, we opted to highlight the interplay of the secondary BH with the mass field.



**Figure A2.** Snapshots of projected gas density (edge-on view) with 1 kpc wide box. In each row, from left to right, there are three different snapshots every 5 Myr of a given simulation. The runs are, from top to bottom, respectively, NAC, CSFSN, and CSFSNBF.





**Figure A3.** Density snapshots for run CBF.